Available on CMS information server **CMS IN 2007/062**

CMS Internal Note

The content of this note is intended for CMS internal use and distribution only

13 November 2007

Performance of the k_t jet algorithm in CMSSW

P. Schieferdecker, M. Vázquez Acosta

CERN, Geneva, Switzerland

A. Oehler, K. Rabbertz

Universität Karlsruhe, Germany

G. Bruno, D. Kçira, L. Quertenmont

Université Catholique de Louvain, Belgium

P. Kurt

University of Cukurova, Turkey

C. Dragoiu, N. Varelas

University of Illinois, Chicago, USA

F. Ratnikov

University of Maryland, College Park, MD, USA

J. Widawsky, M. Zielinski

University of Rochester, Rochester, NY, USA

A. Bhatti

Rockefeller University, USA

G. Dissertori

Institute for Particle Physics, ETH, Zürich, Switzerland

Abstract

First studies of the performance of the k_t jet algorithm and its implementation in CMSSW are summarized. Jet efficiency, response and resolutions in QCD events and the performance of the algorithm in the selection and reconstruction of hadronic decays of weak bosons are presented. The new pile-up and underlying event subtraction method provided in the FastJet package has been tested and first results are also shown.

1 Introduction

The $\boldsymbol{k_t}$ clustering jet algorithm [1] in the longitudinally invariant form is a collinear and infrared safe algorithm, which is suited for hadron colliders.

In this study the jet search is performed in the laboratory frame, using as input calorimeter towers (CaloTowers) or stable hadrons. In the following discussion, $k_{t,i}$ denotes the transverse momentum, y_i the rapidity, ϕ_i the azimuthal angle of object i and D is the distance or radius parameter. For each pair of objects, the quantity

$$d_{ij} = \min(k_{t,i}^2, k_{t,j}^2) \cdot \frac{[(y_i - y_j)^2 + (\phi_i - \phi_j)^2]}{D^2}$$

is calculated. For each individual object, the quantity $d_i = (k_{t,i})^2$ is also calculated. If, of all the values $\{d_{ij}, d_i\}$, d_{kl} is the smallest, then objects k and l are combined into a single new object. If, however, d_k is the smallest, then object k is considered a jet and is excluded from further clustering. The procedure is repeated until all objects were assigned to jets.

In the old CMS reconstruction framework ORCA [2] and within the initial versions of CMSSW [3] up to version 1.2.X, a CMS internal implementation of the $\boldsymbol{k_t}$ algorithm was used. Starting with the 1.3.X CMSSW series, two external implementations were introduced: the library KtJet [4] and a version of the algorithm included in the FastJet package [5]. The $\boldsymbol{k_t}$ implementation of FastJet includes optimizations that make the jet-finding algorithm much faster than the traditional one, i.e. the time spent for the reconstruction of an event is much smaller. The FastJet package also includes additional features such as the pile-up subtraction on an event-by-event basis discussed later in this note. Apart from the $\boldsymbol{k_t}$ algorithm, the FastJet package comprises interfaces to other jet-finding algorithms not used in the studies presented here.

Figure 1 shows the time distribution spent per event by the standard $\boldsymbol{k_t}$ algorithm and the FastJet implementation of it. The samples used for this comparison are standard QCD events generated for validation purposes without and with addition of in-time pile-up contribution for a low-luminosity scenario, $\mathcal{L} = 2 \cdot 10^{33}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. The calorimeter towers (CaloTowers) were used as input and the D-parameter (sometimes also called R-parameter) of the algorithm was set to a value of 1. The FastJet implementation shows a substantial improvement in speed. Also shown is the timing of the simple Iterative Cone algorithm which is used in the trigger at CMS. The FastJet implementation of the $\boldsymbol{k_t}$ algorithm has a comparable speed to the Iterative Cone algorithm and is less sensitive to the CaloTower input E_t threshold applied on top of the noise suppression cuts (Scheme B thresholds [6] have been applied in all cases).

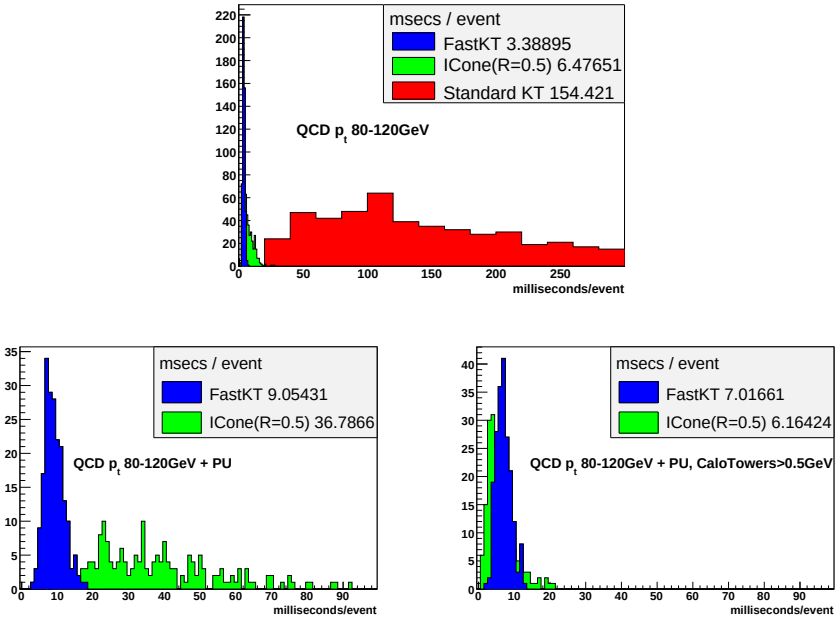

Figure 1: **Top plot:** Distribution of time spent per event by the KtJet implementation of $\boldsymbol{k_t}$ (red), the FastJet implementation of the $\boldsymbol{k_t}$ (blue) algorithm and the Iterative Cone (green) for calorimeter jets. **Bottom plots:** Comparison of the timing distributions of the FastJet implementation of the $\boldsymbol{k_t}$ algorithm and the Iterative Cone with (left) no cut and (right) with a cut of $E_t > 0.5$ GeV on the input calorimeter towers.

2 k_t Algorithm validation in CMSSW

In this section, the output of two different implementations of the k_t algorithm, the KtJet and FastJet packages, is compared for a standard setting of parameters. Ideally, the results should be identical, any observed difference will come from details of the implementation. Since in future versions of the CMS software only the FastJet package will be used, this comparison is crucial to assure a clean transition.

2.1 Analysis Procedure and Event Selection

Standard QCD dijet events were generated with PYTHIA [8] and simulated with CMSSW 1.3.1. In each case two different inputs were used: the four-vectors of the generated stable particles in the event and the four-vectors of the calorimeter towers after the detector simulation. Two output jet-collections were produced which are respectively called generator level jets (GenJets, also particle jets or hadron jets) and calorimeter jets (CaloJets).

For both KtJet and FastJet the following parameters were used:

- the D-Parameter is set to 1
- all stable particles are taken as input for reconstructing GenJets
- Calo towers with $E_T > 0.5$ GeV are used as input for the reconstruction of CaloJets
- the output was restricted to Jets with $p_{\mathrm{T}} > 1.0$ GeV.

The distribution of the transverse momentum of jets produced from the Monte Carlo (MC) samples is shown in Fig. 2. The events are generated within the two ranges of $50 - 80$ GeV and $800 - 1000$ GeV of the transverse momentum $\hat{p}_T$ of the partons participating in the hard interaction.

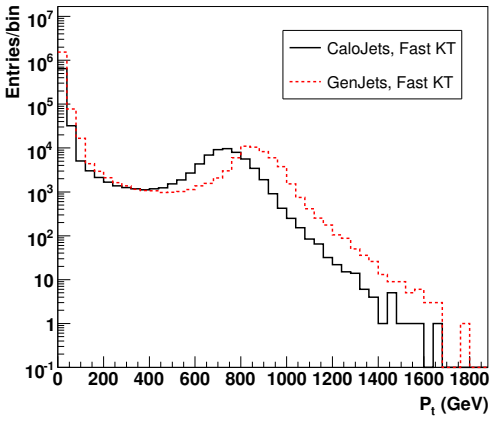

Figure 2: Calorimeter and generator jet p_{T} distribution from the FastJet package.

2.2 Matching Results

The closest jets of both collections have been matched according to their distance in

$$\Delta R = \sqrt{[\Delta\phi]^2 + [\Delta y]^2}, \tag{1}$$

where $\Delta\phi$ and Δy are the differences of the four-vectors in the azimuthal angle and the rapidity, respectively. When the four-vectors from stable particles of generated events, which are known at the level of double precision i.e. up to 15 decimals, are used as input of both implementations of k_t, no difference in the jet output can be observed within the selected events. This means that ΔR and the difference in p_{T} of matched jets is equal to zero to this order of precision. Additionally, every jet of every event has a match in the other collection.

When applied on calorimeter towers which are stored in single precision in the event files, there is a slight difference in the output jet-collections of FastJet and KtJet. About $1.5‰$ of the jets do not match on the level of double precision as listed in Table 1. In Fig. 3 the relative difference in transverse momentum of all jets is shown to

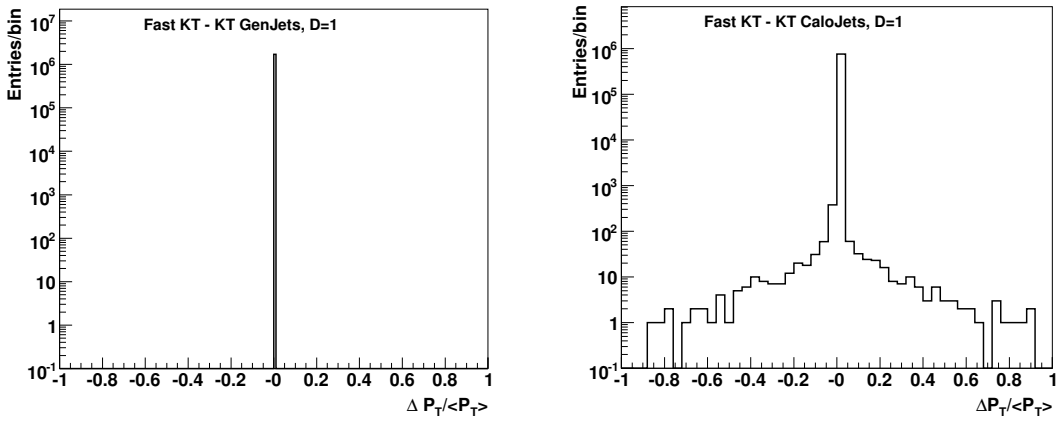

Figure 3: Relative difference of the jet transverse momenta of the matched jets from FastJet and the KtJet packages for **Left plot:** generator input in double precision and **Right plot:** calorimeter input in single precision.

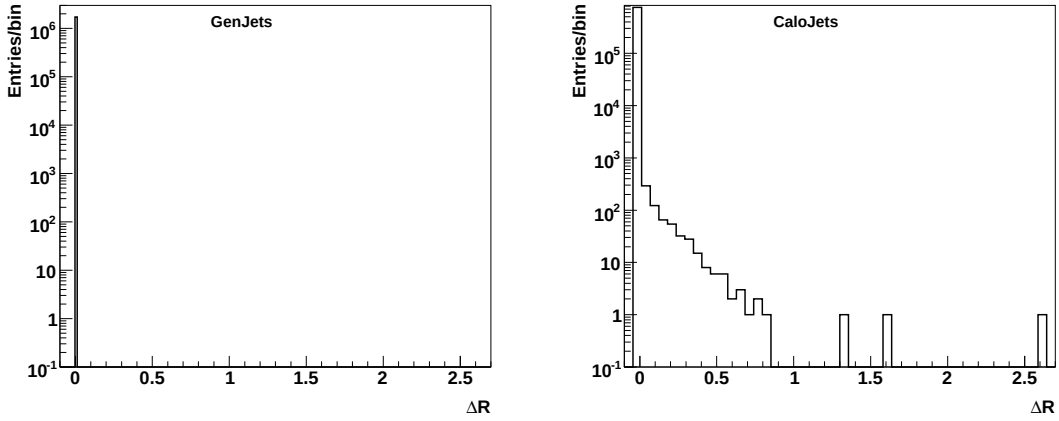

Figure 4: Difference in ΔR of FastJet and KtJet matched jets for **Left plot:** generator input in double precision and **Right plot:** calorimeter input in single precision.

differ with an RMS of $6.34 \cdot 10^{-3}$. The right plot in Fig. 4 shows the difference in ΔR, which shows an RMS of $6.57 \cdot 10^{-3}$.

Figure 5 (right) shows the p_{T} of all jets which have a different p_{T} on the level of double precision. This distribution is similar to the p_{T} distribution of all jets, as is shown in Fig. 2. The relative differences in p_{T} are small in most cases, which indicates that only occasionally constituents become clustered into different jets.

2.3 Differences in the Implementation

One difference of both algorithms is the different treatment of four-vectors with negative mass squared in case of the rapidity calculation, as well as a different way of calculating the rapidity numerically. The FastJet package has a special four-vector implementation called PseudoJet, which checks the created four-vectors during the initialisation step and fixes the mass squared to be positive or zero. Using

$$k_t^2 = P_x^2 + P_y^2 \tag{2}$$

Table 1: Output of both algorithms for matched jets according to the closest distance ΔR

Input Type	Number of Jets	Number of Different Jets	Number of non-matched jets
MC particles	1710285	0	0
calo towers	754869	1135	183

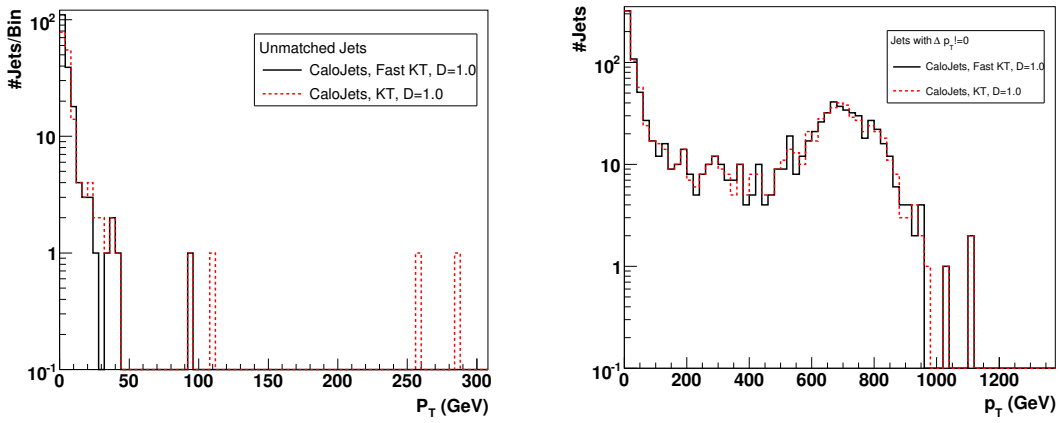

Figure 5: Distributions of jets which are clustered differently in KtJet and FastJet. p_{T} distribution of jets which **Left plot:** have not been taken into account for matching and **Right plot:** do not have equal p_{T} up to double precision.

it sets the rapidity to a very large value of $\mathrm{sign}(P_z) \cdot (10^5 + |P_z|)$ in the rare case of $E = |P_z|$ and $k_t^2 = 0$. The rapidity is calculated as

$$y = \mp \frac{1}{2} \ln \left(\frac{k_t^2 + \max(0, (E + P_z)(E - P_z) - k_t^2)}{(E + P_z)^2} \right). \tag{3}$$

For $P_z > 0$ the minus sign is selected in Eq. 3.

In contrast to that, the KtJet package has a four-vector implementation which derives from a CLHEP [7] four-vector and has some additional features built in. The rapidity calculation takes place within the CLHEP part. Also in this case some properties of the four-vectors are checked. Four-vectors with $|E| < |P_z|$ are not accepted and an exception is thrown. However, four-vectors with $|E| < |P|$, which are not rejected by the first condition, are accepted. In this case the rapidity is calculated using

$$y = \frac{1}{2} \ln \left(\frac{E + P_z}{E - P_z} \right). \tag{4}$$

Both ways of calculating the rapidity have a different approach for handling particles with negative mass squared (so-called tachyonic particles / masses). While KtJet would act like mirroring the mass at zero, FastJet would fix it to zero. In addition they are differently affected by numerical fluctuations.

2.4 Jet Constituents in CMSSW

As an example of problems which can arise, the jet constituents as currently employed in CMSSW are discussed. If the input four-vectors are derived on the fly in double precision, as done in Fig. 6 left for generator particles, negative mass squares do not appear. If, however, the input four-vectors are read in from data files, where they have been written in single precision to reduce storage requirements, the change in numerical precision alone when casting to double precision may already lead to negative mass squares, see Fig. 6 right for example for calorimeter towers. Consequently, discrepancies in jet reconstruction between the two discussed implementations can be observed.

2.5 Rapidity Calculation

Figure 7 shows the difference in rapidity when either the FastJet or the KtJet method is applied to calculate the rapidity of the same input four-vector. For towers with tachyonic masses the calculation leads to remarkable differences in rapidity on the scale of $\pm 10^{-3}$, which leads to a different clustering of towers. Despite these major differences, also for $\frac{E-|P|}{E} > 0$ small differences of the order of $\pm 10^{-6}$ can be seen, which also can lead to a different clustering, but with less total effect than for the $\frac{E-|P|}{E} < 0$ case. The same holds true for the input of GenJets, where $\frac{E-|P|}{E} < 0$ occurs.

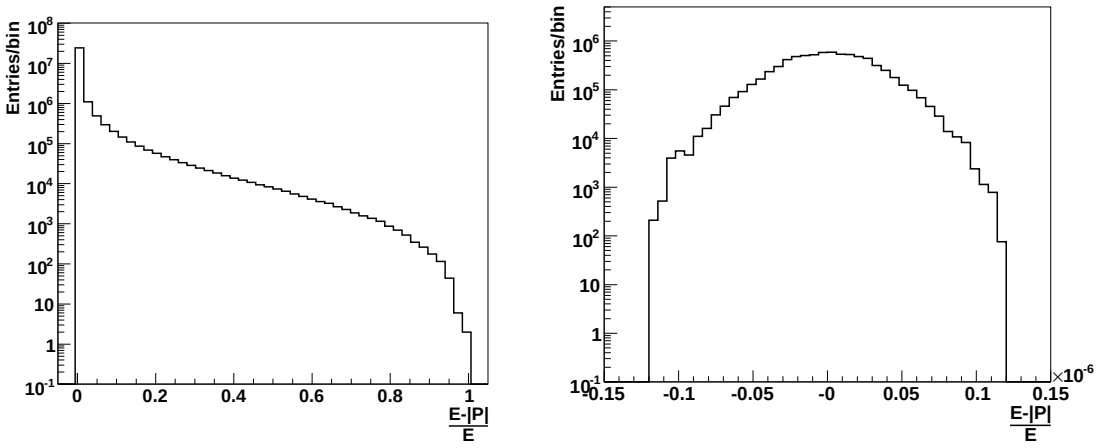

Figure 6: $\frac{E-|P|}{E}$ distribution for the example of **Left plot:** generator particles in double precision and **Right plot:** Calorimeter Towers read from storage in single precision.

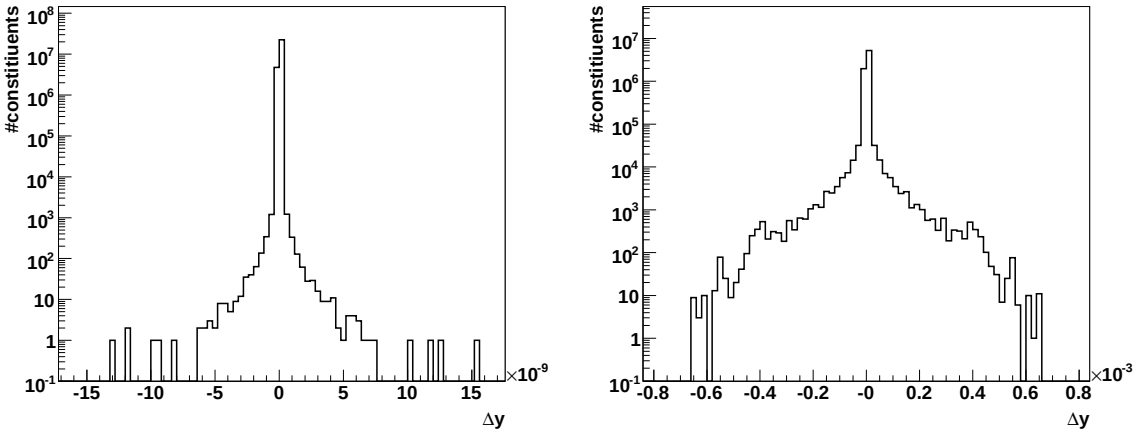

Figure 7: Difference of the rapidity calculation in FastJet and KtJet for **Left plot:** generator input in double precision and **Right plot:** calorimeter input in single precision.

2.6 Fixed Input Collections

If one fixes the jet input for input four-vectors with $E - |P| \leq 0$ to $E = |P| + \epsilon$, with

$$\epsilon = \frac{1}{100000} + \frac{|E|}{100000} \tag{5}$$

the difference observed within the matching procedure is strongly reduced, as can be seen from Table 2. In this case the algorithms have produced exactly the same number of jets for all events and only a very small fraction of the matched jets differ in p_{T}.

In this section the effect of tachyonic masses of the input four-vectors has been shown to contribute to a different clustering of both $\boldsymbol{k_t}$ implementations. The fixed outputs are not identical, which arises from the different numerical behaviour of both rapidity calculation methods. This only is observable when calorimeter input is used, where numerous inputs have exactly equal coordinates due to the fact that towers ϕ and η coordinates are derived from the geometrical arrangement of the objects within the detector. Thus small differences in this value can lead to a different value of the resolution parameter, thus clustering an input object into a different jet. Despite the case of ill defined input vectors, both implementations of the $\boldsymbol{k_t}$ algorithms have shown to produce the same collection of jets.

Table 2: The jet output of both algorithms is matched according to closest ΔR-Distance for corrected jet input

Input Type	Number of Jets	Number of Different Jets	Number of non-matched Jets
MC particles (double precision)	1704270	0	0
CaloTowers (single precision)	754773	3	0

3 Jet Efficiency, Response and Resolutions

The response of calorimeter jets reconstructed with any jet clustering algorithm is defined with respect to the hadron-level transverse energy $E_{\rm T}^{\rm gen}$ of the jet as

$$\text{jet response} = \frac{E_T^{\rm calo}}{E_T^{\rm gen}} \ . \tag{6}$$

$E_{\rm T}^{\rm gen}$ is hereby derived for each jet algorithm by clustering all stable generated particles in the event, producing a collection of reference jets (referred to as "generated jets", "particle-level jets", or "hadron-level jets"). The jet response is measured using a QCD dijet MC sample generated with the PYTHIA event generator and processed with the full CMS detector simulation (CMSSW 1.2.3) and event reconstruction (CMSSW 1.3.1). The same sample was used to derive the official CMSSW 1.3.1 jet energy corrections (JES) [9].

Only calorimeter (reconstructed) jets which can be associated with a generated jet in η-ϕ space through

$$\Delta R = \sqrt{\left[\Delta\eta\right]^2 + \left[\Delta\phi\right]^2} < 0.3 \tag{7}$$

are considered. The ΔR distribution for the studied jet algorithms is shown in Fig. 8 for two D parameter values of 0.6 and 1 in different $p_{\rm T}$ regions. The matching between the calorimeter jets and the generated jets improves substantially as $p_{\rm T}$ increases.

Figure 9 shows the reconstruction efficiency as a function of generated jet $p_{\rm T}$ in the barrel- ($|\eta| < 1.4$), endcap- ($1.4 < |\eta| < 3.0$), and forward- ($3.0 < |\eta| < 5.0$) regions. The jets reconstructed with D=0.6 have a faster turn-on in the reconstruction efficiency as a function of $p_T^{\rm gen}$. Figure 10 shows the reconstruction efficiency as a function of generated jet $p_{\rm T}$ for a relaxed matching criteria of $\Delta R < 0.5$. Figures 11 shows the reconstruction efficiencies as a function of η in different $p_T^{\rm gen}$ regions ([15,25], [40,50], [100,300] GeV). The reconstruction efficiency improves as $p_{\rm T}$ increases. Figure 12 shows the reconstruction efficiency as a function of η for a relaxed matching criteria of $\Delta R < 0.5$. The reconstruction efficiency $\epsilon_{\rm RECO}$ is hereby defined as the ratio of the number of generated jets that matched a calorimeter jet to the total number of generated jets.

To measure the response as a function of $E_{\rm T}^{\rm gen}$, the sample is divided in bins of $E_{\rm T}^{\rm gen}$. The jet response distribution $E_T^{\rm calo}/E_{\rm T}^{\rm gen}$ in each bin is fitted with a Gaussian. The mean of the fit is considered as the jet response of the bin, while the width of the Gaussian is interpreted as the absolute jet resolution $\sigma(E_T^{\rm calo}/E_{\rm T}^{\rm gen})$. The jet response as a function of $E_{\rm T}^{\rm gen}$ is finally derived as the fitted jet response versus the mean of the $E_{\rm T}^{\rm gen}$ distribution in each bin. As the jet response distributions show sizeable asymmetric non-Gaussian tails at low $p_{\rm T}$, the fit range is limited to the mean $\pm 1.5 \cdot RMS$ of the distribution. Figure 13 shows the resulting response for jets reconstructed with the FastJet implementation of the $\boldsymbol{k_t}$ algorithm. The response is computed before (left) and after (right) the official CMSSW 1.3.1 jet energy corrections are applied. Jets generated in the barrel, endcap, and forward regions of the calorimeter are treated separately. The jet response obtained with a D parameter value of 0.6 is similar to the one obtained with a value of 1.

The relative energy resolution

$$\frac{\sigma(E_T^{\rm calo}/E_T^{gen})}{E_T^{\rm calo}/E_T^{gen}} \tag{8}$$

is shown in Fig. 14. The individual points are fitted with the function

$$\sqrt{\left(\frac{a}{E_T^{gen}}\right)^2 + \left(\frac{b}{\sqrt{E_T^{gen}}}\right)^2 + c^2} \tag{9}$$

and the extracted resolution parameters are given in the legend of each figure. The energy resolution is slightly improved with a D parameter values of 0.6 compared to 1. The η and ϕ position resolutions are shown in Fig. 15 and 16. The position resolution improves substantially with a D parameter values of 0.6 compared to 1.

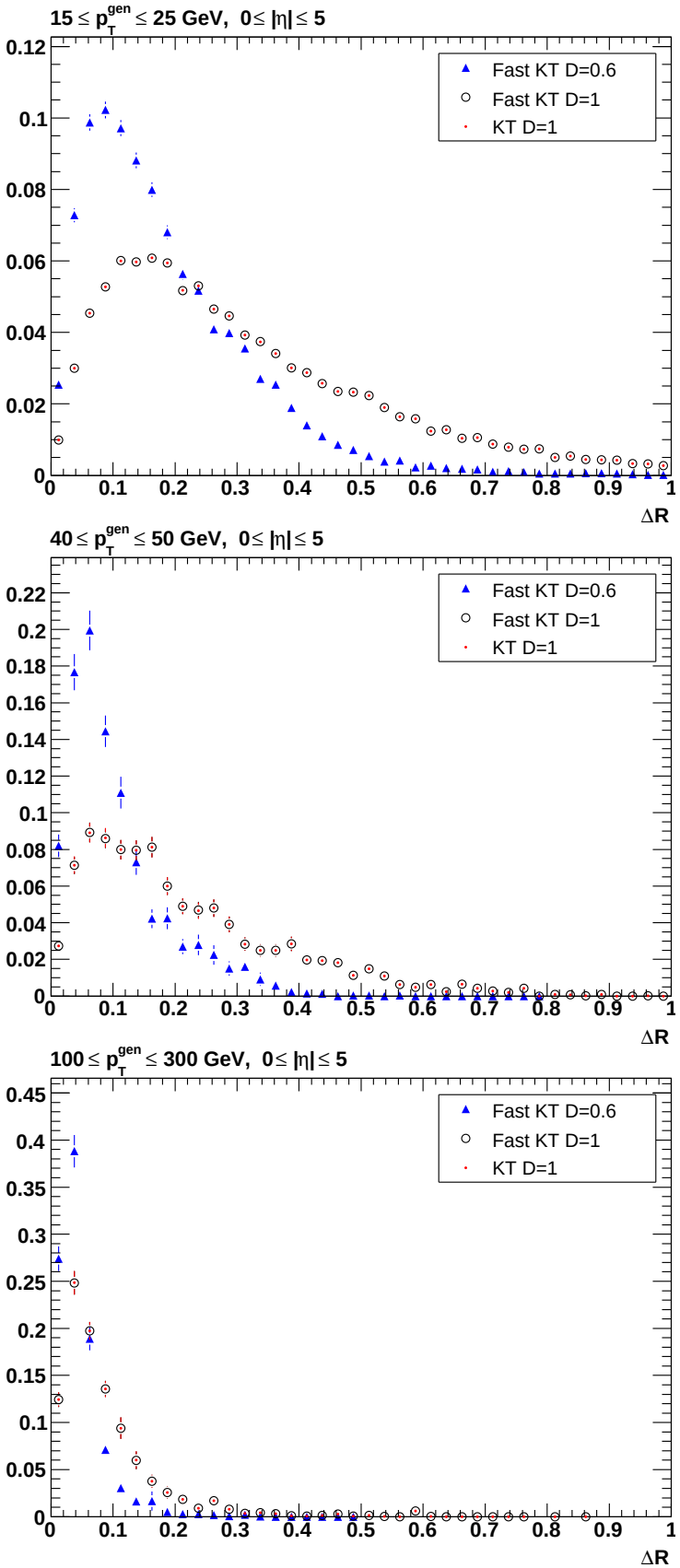

Figure 8: ΔR distribution for different $\boldsymbol{k_t}$ algorithm D parameter values of D=0.6 and D=1 in different p_{T} regions. Results using the FastJet and KtJet package are also compared.

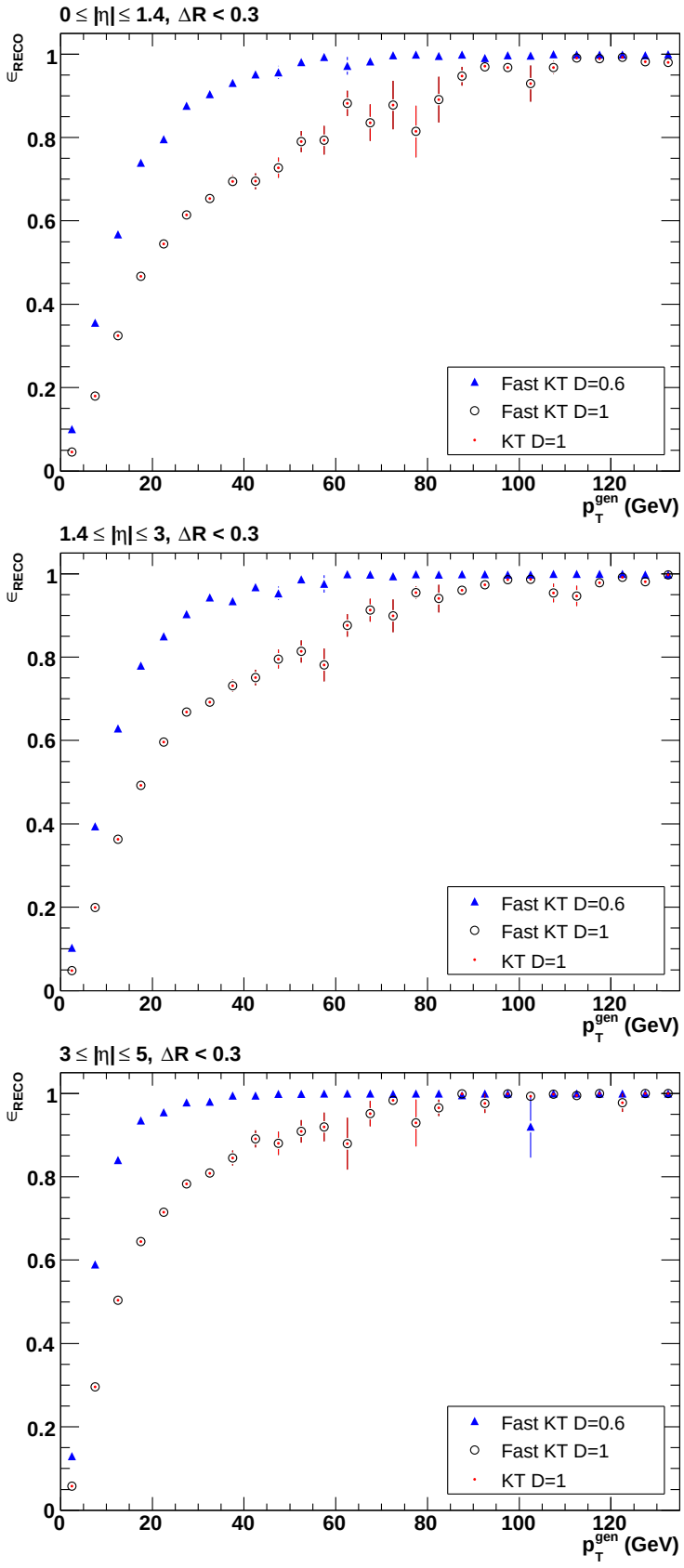

Figure 9: Reconstruction efficiency as a function of jet p_{T} in the barrel (top), endcap (center), and forward (bottom) regions are shown seperately for $\Delta R < 0.3$.

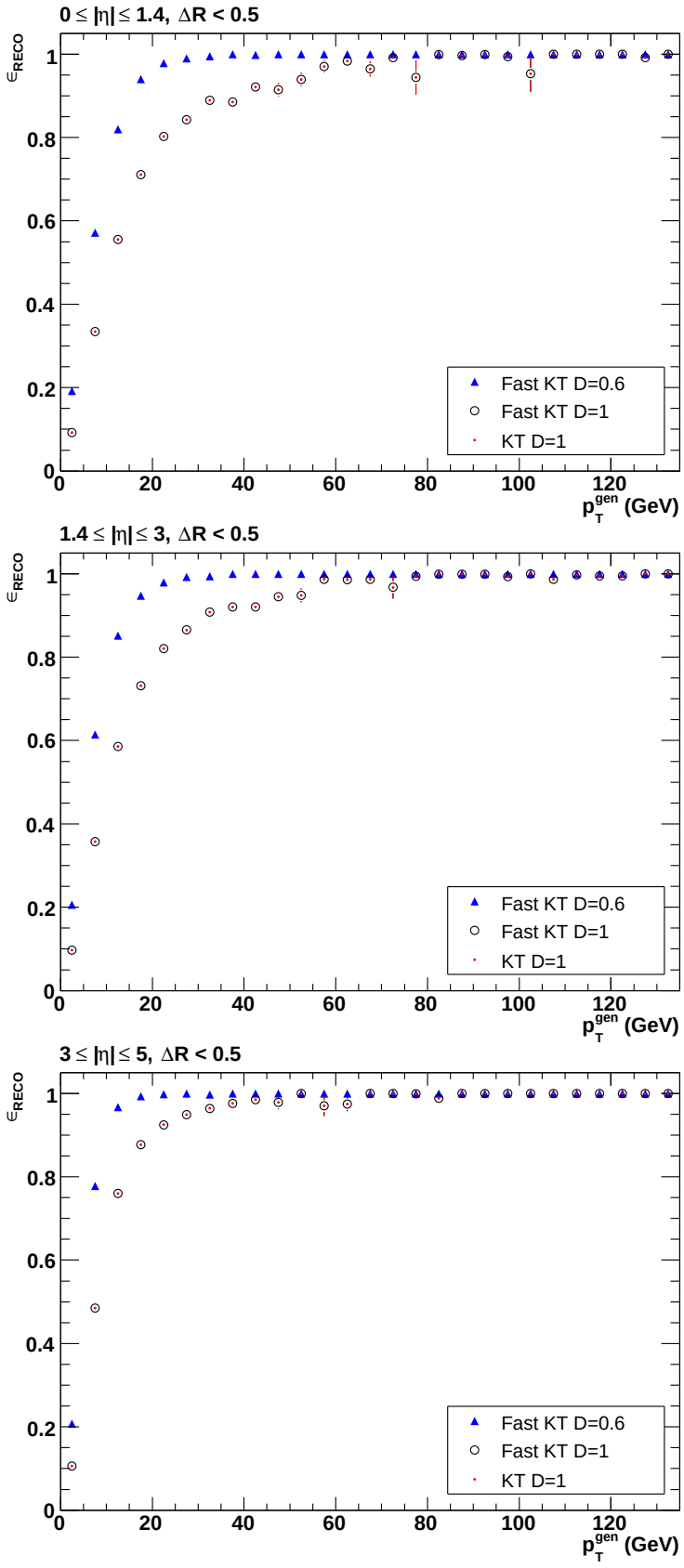

Figure 10: Reconstruction efficiency as a function of jet p_{T} in the barrel (top), endcap (center), and forward (bottom) regions for $\Delta R < 0.5$.

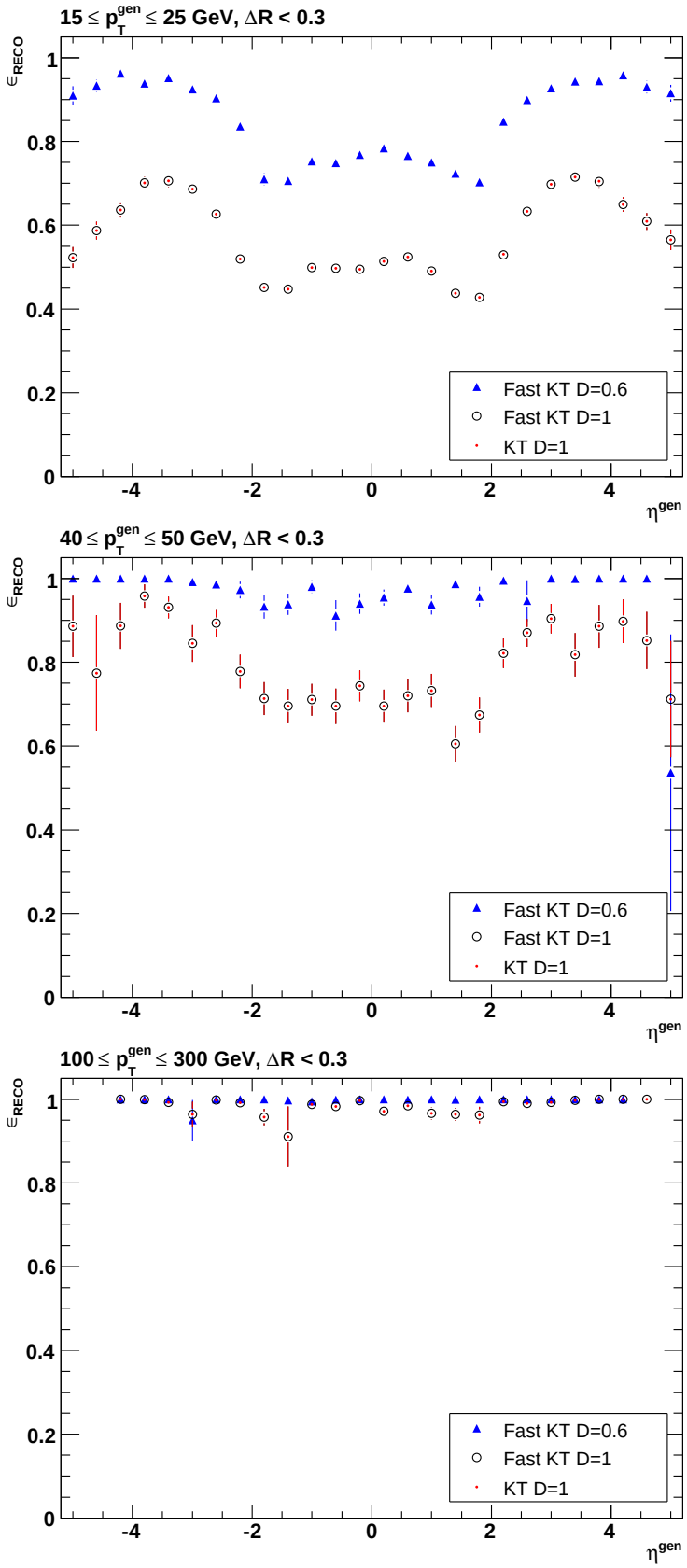

Figure 11: Reconstruction efficiency as a function of the jet η in different jet p_{T} ranges for $\Delta R < 0.3$.

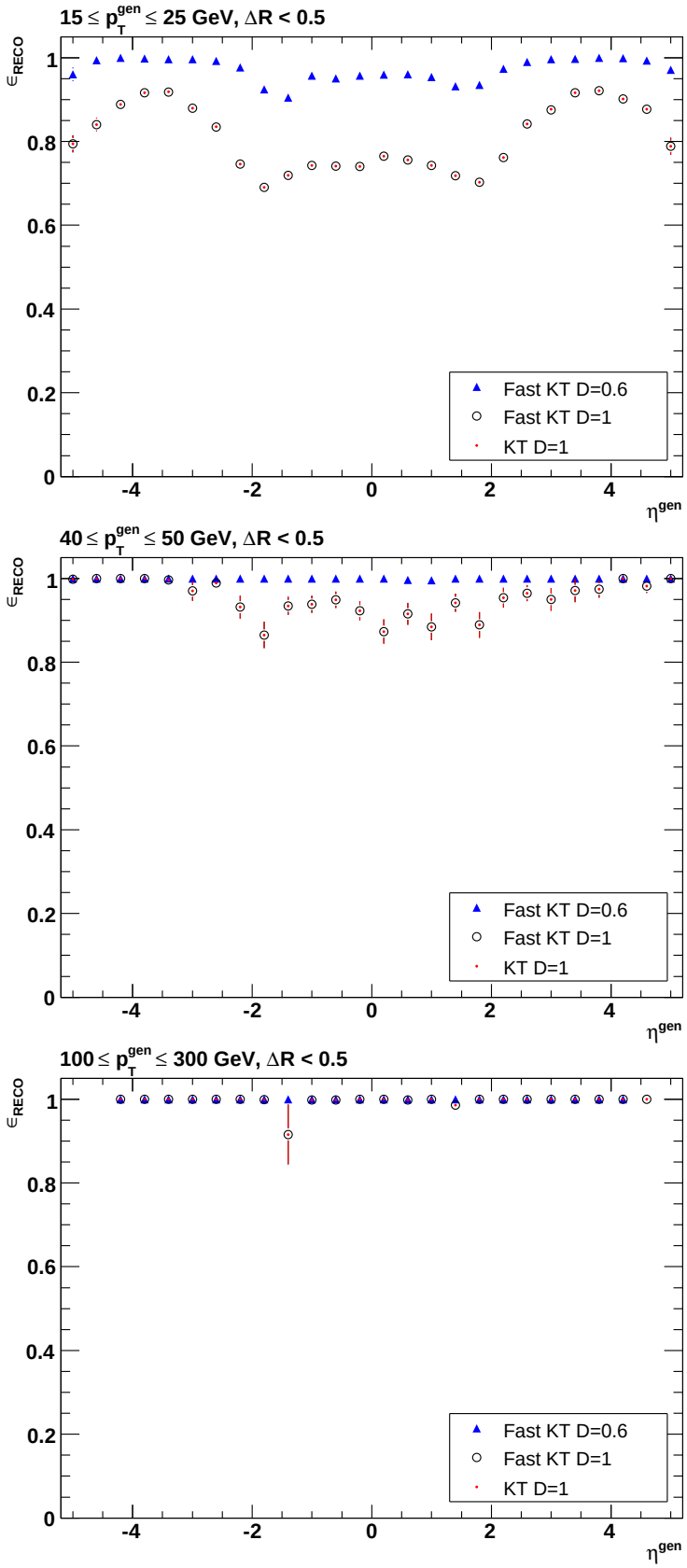

Figure 12: Reconstruction efficiency as a function of the jet η in different jet p_{T} ranges for $\Delta R < 0.5$.

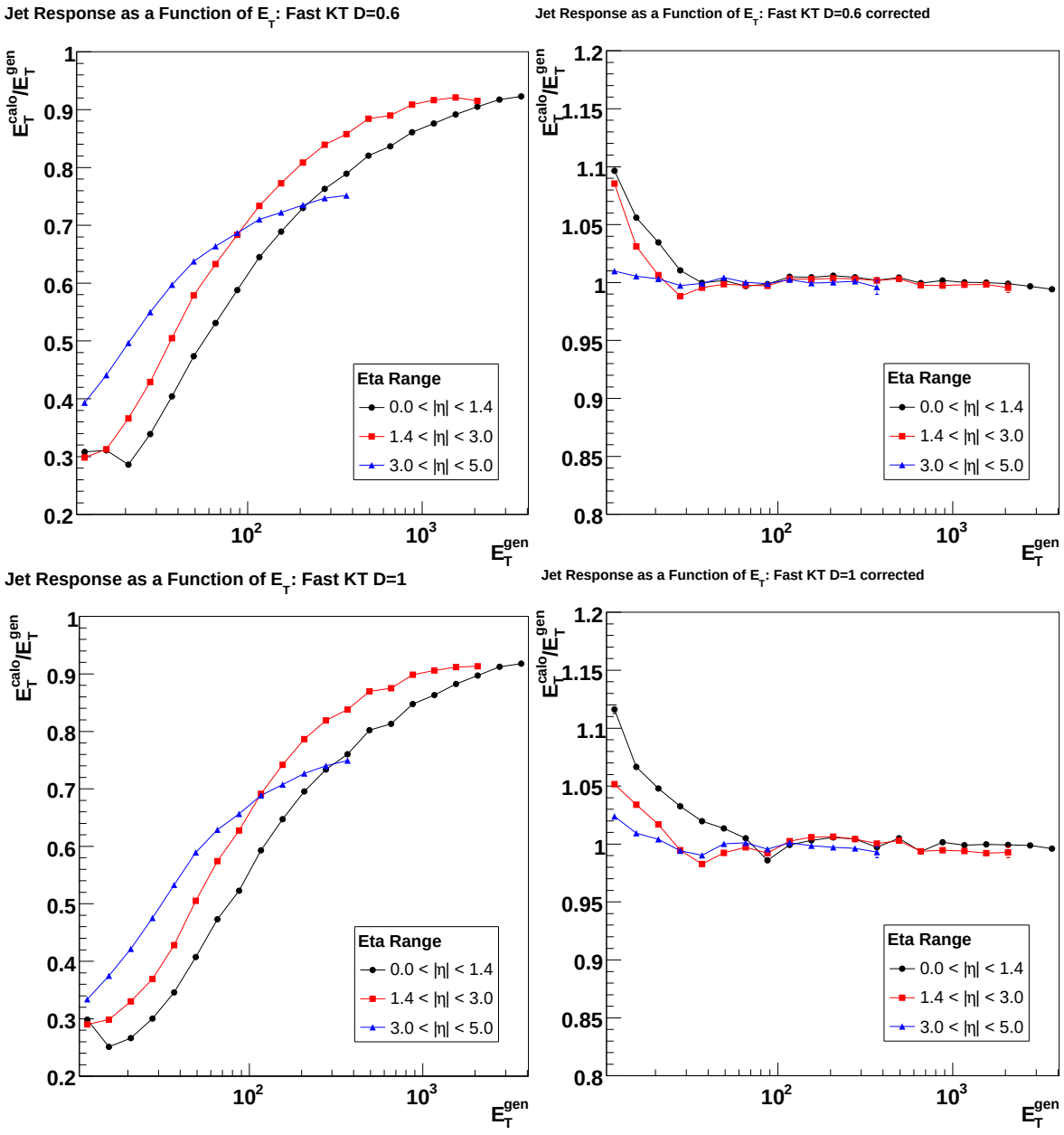

Figure 13: Jet response $E_T/E_{\mathrm{T}}^{\mathrm{gen}}$ as a function of generated jet E_T for jets reconstructed with the FastJet implementation of the k_t algorithm with D=0.6 (top) and D=1 (bottom) before (left) and after (right) application of the official CMSSW 1.3.1 jet energy corrections.

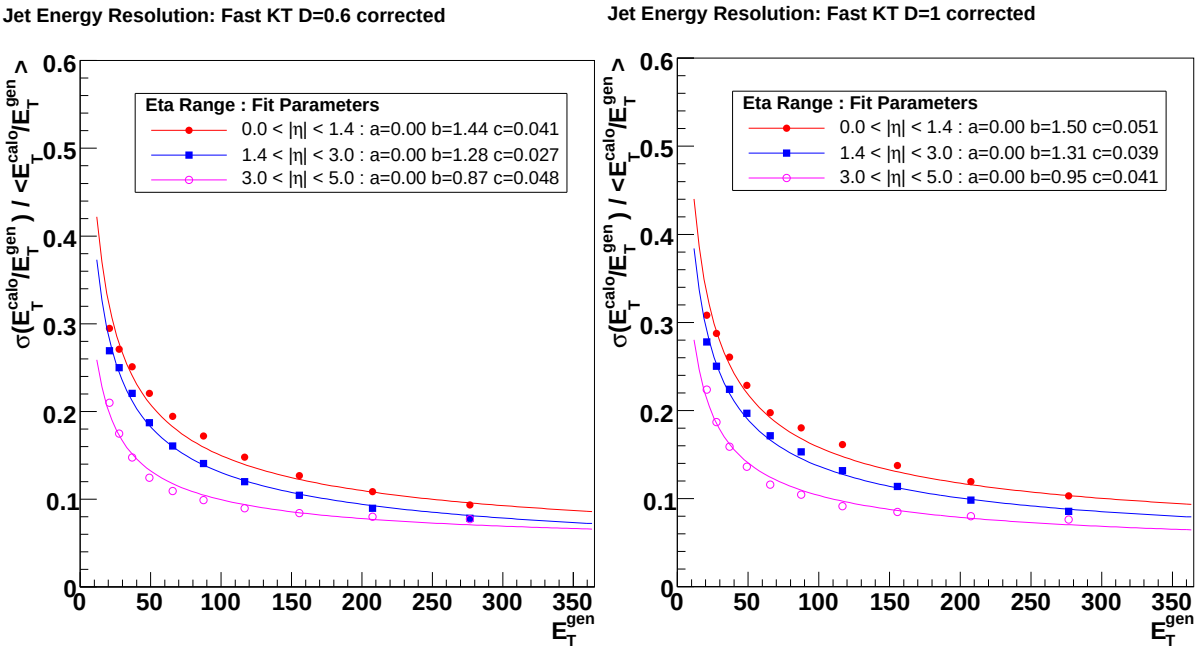

Figure 14: Jet energy resolutions as a function of generated jet E_T for jets reconstructed with the fast k_t algorithm with **Left plot:** D=0.6 and **Right plot:** D=1. The standard CMSSW 1.3.1 jet energy corrections are applied before the resolution is extracted.

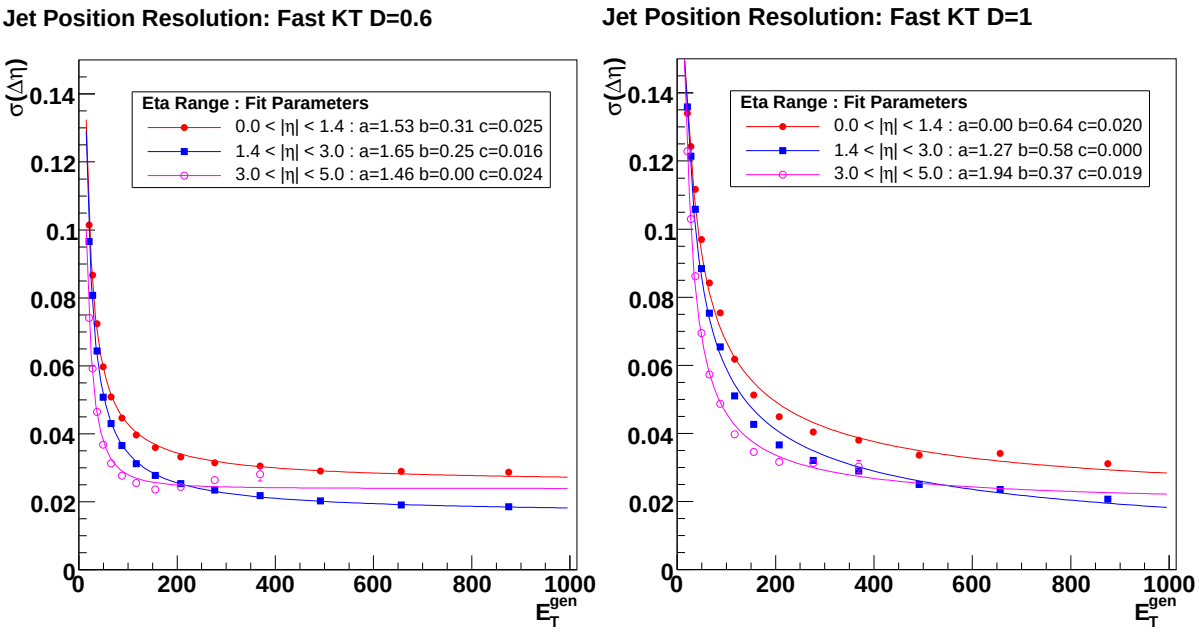

Figure 15: η-resolutions as a function of generated jet E_T for jets reconstructed with the FastJet implementation of the k_t algorithm with **Left plot:** D=0.6 and **Right plot:** D=1.

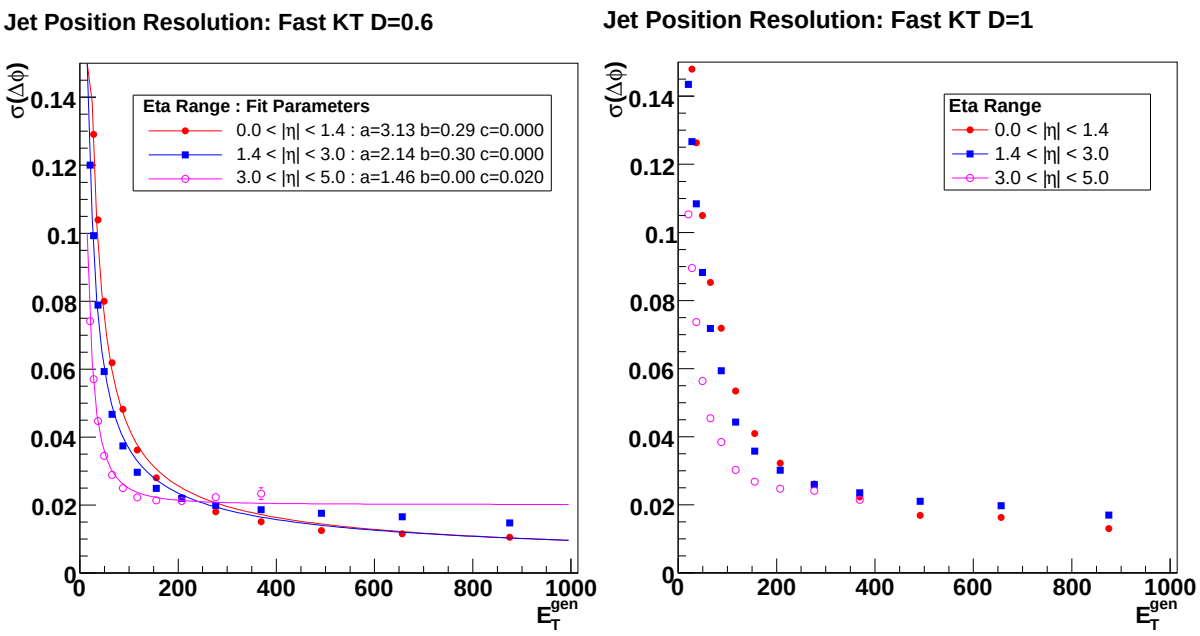

Figure 16: ϕ-resolutions as a function of generated jet E_T for jets reconstructed with the FastJet implementation of the k_t algorithm with **Left plot:** D=0.6 and **Right plot:** D=1 (fits are not shown due to their bad convergence).

4 Energy Distribution within a Jet

Although in the k_t jet clustering algorithm the objects (particles, towers, tracks or partons) are combined based on their relative transverse momenta, the resulting jet is mainly localized in $y - \phi$ space. In other words, although k_t clustered jets do not have a precise boundary in physical space, the jet properties are similar to the properties of jets clustered by the cone clustering algorithms. This can be seen by studying the energy distribution within a jet which is characterized by the differential jet shape, $\rho(r)$, defined as

$$\rho(r) = \frac{\sum P_T(r - \Delta r/2, r + \Delta r/2)}{\Delta r \sum P_T^{Jet}}$$

where the sum is over all the jet constituents in the range $(r - \Delta r/2, r + \Delta r/2)$ in the numerator where $r = \sqrt{(y_{jet} - y_c)^2 + (\phi_{jet} - \phi_c)^2}$ with (y_{jet}, ϕ_{jet}) and (y_c, ϕ_c) being the position of the jet and the constituents. The denominator $\sum P_T^{Jet}$ is the scalar sum of the transverse momenta of all the jet constituents.

Figure 17 shows the jet shape distribution $\rho(r)$ as a function of the distance from the jet centroid for k_T-jets from a dijet MC sample generated and simulated using CMSSW 1.3.1. The particles and calorimeter towers belonging to the reconstructed jets are used for jets with $50 < P_T^{Jet} < 80$ GeV. As expected, in contrast to the cone jets, k_T jets extend to a larger area in physical space and do not have a sharp cut off. However, most of the energy is concentrated at the core of the jets. The particle jets include the particles down to very small momenta and thus have large area. The calorimeter jets are reconstructed from towers which have minimum $E_T > 0.5$ GeV and pass scheme B thresholds. These threshold requirements limit the area of the jets. In addition, charged low energy particles do not reach the surface of the ECAL due to the large magnetic field in CMS.

The differential jet shape of the particle jets is compared with the calorimeter jets in Fig. 18 for D parameter values of 0.4, 0.6 and 1.0. The particle jets extend to a larger physical area than calorimeter jets. Quantitatively, 90% of the jet energy is confined to $r < 0.25, 0.37$ and 0.63 for $D = 0.4, 0.6$ and 1.0 particle jets while for calorimeter jets the values are $r < 0.27, 0.37$ and 0.55, respectively.

The fraction of jet transverse energy outside $r = 0.5$ is 1%, 5% and 15% for $D = 0.4, 0.6$ and 1.0 respectively for particle jets which can be compared to 1%, 4% and 12% for the calorimeter jets.

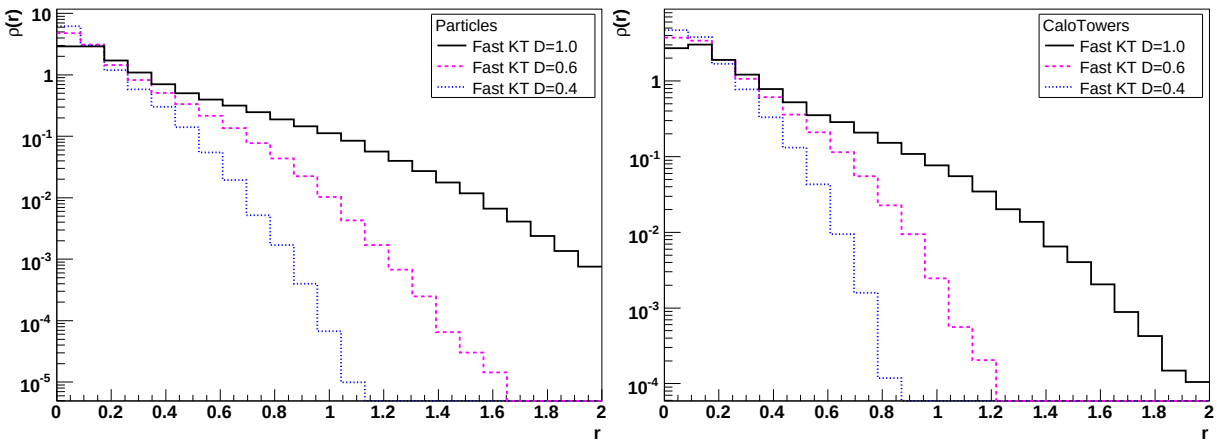

Figure 17: Normalized transverse energy density distributions, $\rho(r)$ for particle (left) and calorimeter (right) k_T jets reconstructed using the clustering parameter $D = 0.4$, 0.6 and 1.0. As expected, the jets occupy larger area as the D parameters increases.

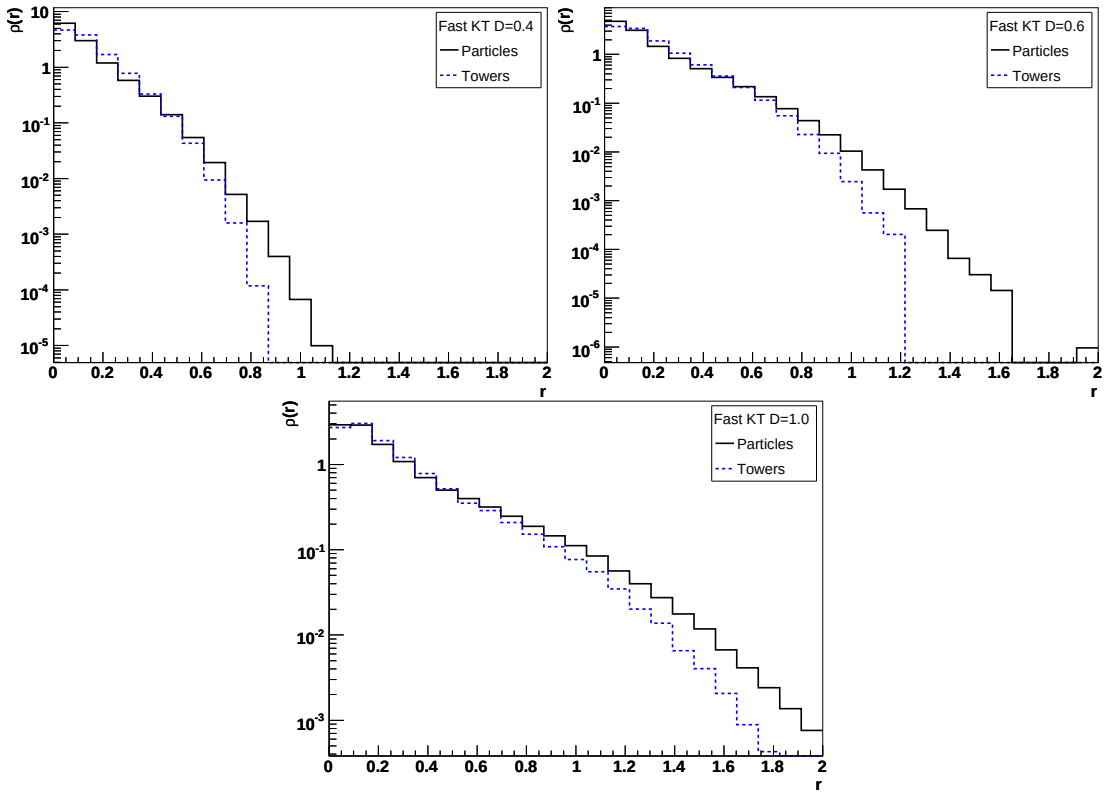

Figure 18: Normalized transverse energy density distributions ($\rho(r)$) in k_T jets reconstructed using the clustering parameter D=0.4, 0.6 and 1.0. The particle jets (solid line) extend much further than the calorimeter jets (dotted line).

5 k_t Performance in Reconstruction of Hadronic Decays of Weak Bosons in Different Energy Regimes

In this section studies are presented of the performance and robustness of the k_t jet-finding algorithm for hadronic decays of standard model weak bosons and of a possible new Z$'$ boson at the TeV scale. These benchmark decays serve to test the different regimes of jet reconstruction at CMS: the low energy (Z boson), more complex environment involving many jets (W boson from top pair production) and higher energy collimated jets (Z$'$).

5.1 Samples Used

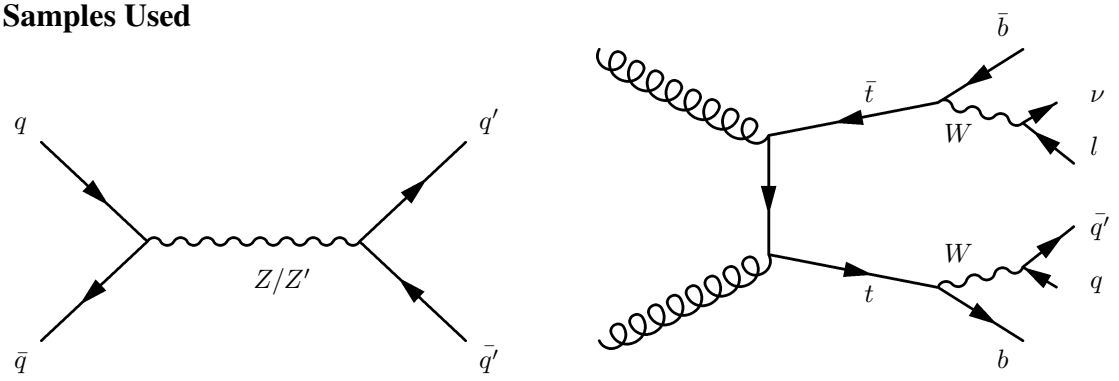

Figure 19: **Left plot:** Diagram of Z or Z$'$ decay into two quarks. **Right plot:** Diagram of top pair production from gluon-gluon fusion with the lower W decaying into two quarks and the upper one into two leptons.

The following event samples were generated at the parton level using the MadGraph [10] matrix element generator:

- Hadronically decaying Z bosons: $Z \rightarrow q\bar{q}$ with mass $m_Z = 91.18$ GeV, 7000 events.
- Hadronically decaying Z$'$ bosons: $Z' \rightarrow q\bar{q}$ with mass $m_{Z'} = 700$ GeV, 5000 events.
- Semileptonic top pair decays: $t\bar{t} \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}l\nu q\bar{q}$, 9150 events corresponding to a total luminosity of 25 pb^{-1}. The mass of the W used for the generation was 80.4 GeV.

These parton level events were interfaced to the Pythia Monte Carlo simulation, version 6.227, that performed the parton showering and hadronization. This was done using the MadGraph interface package in CMSSW 1.3.1. The underlying event model Tune A was used. For all the above samples, events were generated with and without inclusion of in-time pile-up events in the low-luminosity scenario, $\mathcal{L} = 2 \cdot 10^{33}$ cm^{-2}s^{-1}, with a mean of 5 interactions per event. The schematic diagrams for these samples are shown in Fig. 19

5.2 k_t Setting and Inputs

The implementation of the k_t jet-finder in CMSSW1.3.1 is the one from the FastJet package version 2.0. Six different values of the "radius" parameter were used in this study, $D = 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$.

For jet finding at the generator hadron level the input consists of all stable particles (decay length larger than 10 mm), including neutrinos. No cut is applied on the transverse energy of the input particles at this level.

At the detector level three different inputs were used, so three cases are distinguished:

1. Calorimeter: the calorimeter towers above a transverse energy of 0.5 GeV are used as input to the jet reconstruction.
2. Tracking: charged tracks reconstructed using the Kalman filter (so called CTFWithMaterialTracks) with $p_T > 0.9$ GeV are used as input.
3. Particle Flow: the particle flow (PFLOW) algorithms in CMSSW 1.3.1 are run and PFLOW candidates are reconstructed. The jet-finding is then run over these PFLOW objects.

From now on the jets resulting from the above reconstruction will be referred to as Hadron (Particle) Jets, Calo Jets, Track Jets, and PFLOW Jets, respectively.

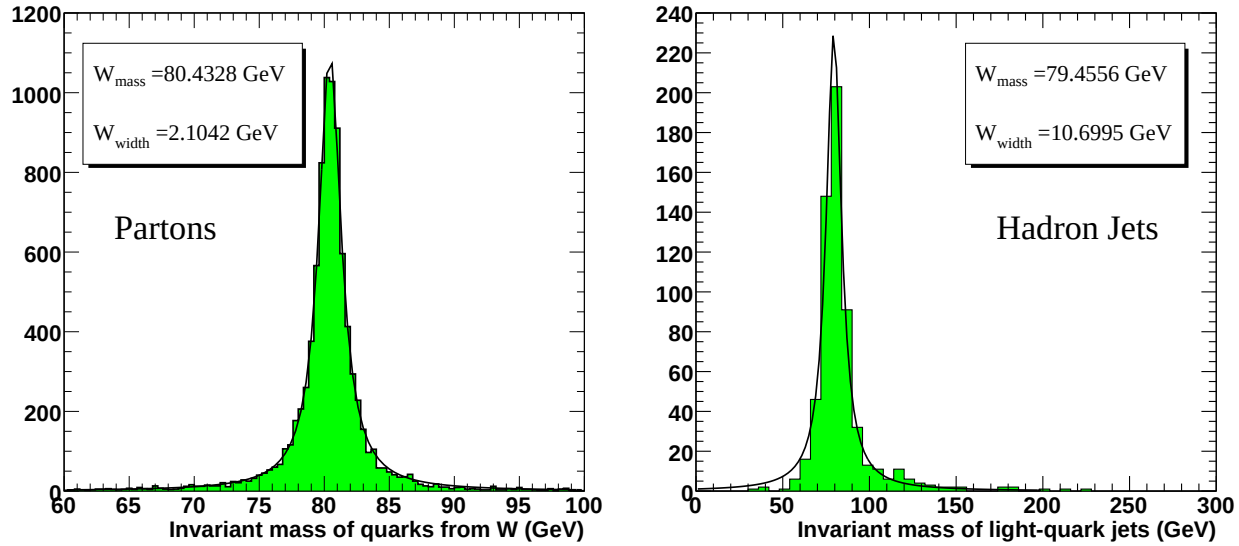

Figure 20: W mass as reconstructed from the partons (left) and from hadron level jets with D=0.4 (right). The plots include the Breit-Wigner fit used to extract the mass and the width.

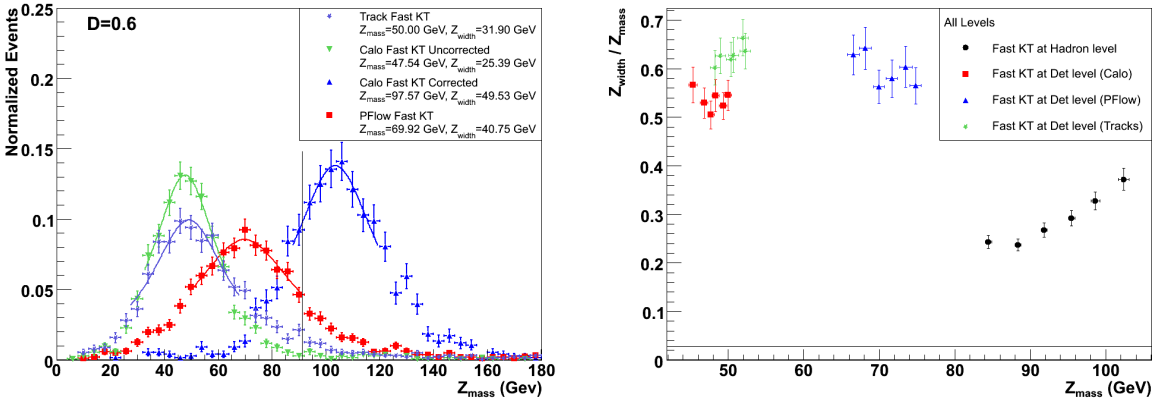

Figure 21: **Left plot:** Reconstructed Z mass at the detector level using $\boldsymbol{k_t}$ jets with D=0.6 for uncorrected and corrected (blue triangles) jets. The Breit-Wigner fit and the extracted values for the mass and the width are also shown. **Right plot:** Width / mass of the Z at the detector and hadron level for different inputs and increasing D-parameter values of the $\boldsymbol{k_t}$ jet-finding algorithm in the range 0.4 to 0.9.

5.3 Event Selection

For the Z and Z$^{'}$ analysis, events were selected with at least two hard jets: $|\eta_{\rm jet}| < 2.4$, $p_{T,\rm jet} > 10$ GeV. The jets were accepted if matched with the originated quarks from the boson decay within a ΔR in the (η, ϕ) plane smaller than 0.3 units. The efficiency of this matching criterion was calculated. The matched jets are used for the reconstruction of the Z/Z$^{'}$ mass.

In the analysis of hadronic decays of W in top quark pair production, events were selected with the following conditions:

- Have a single isolated lepton
  - if the lepton is an electron: $p_T > 26$ GeV and $|\eta_{\rm electron}| < 2.4$
  - if the lepton is a muon: $p_T > 20$ GeV and $|\eta_\mu| < 2.1$
  - for a lepton to be isolated: $|\mathcal{I}(R = 0.2)| < 0.1$, where the isolation variable $\mathcal{I}$ is defined as the relative difference between the transverse momentum of the lepton and the sum of transverse momenta of the charged tracks within a cone of radius $R = 0.2$ from the lepton in the (η, ϕ) plane.

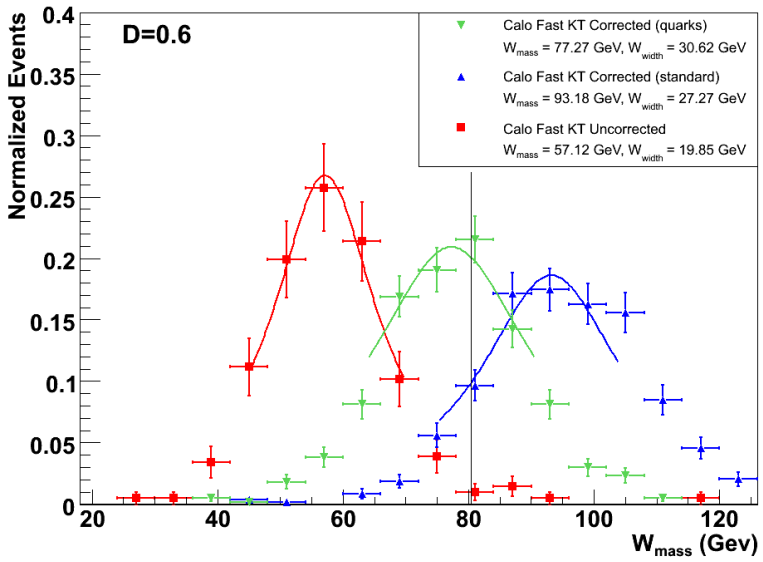

Figure 22: Reconstructed W mass at the calorimeter level using k_t jets with D=0.6. Points using calorimeter cells as well as points with two different jet energy scale corrections are shown. The Breit-Wigner fit and the extracted values for the mass and the width are included for each set of points.

- Have four hard jets after the removal of the jet corresponding to the isolated lepton. The jets must pass the cuts: $|\eta_{\rm jet}| < 2.4$, $p_{T,\rm jet} > 25$ GeV.

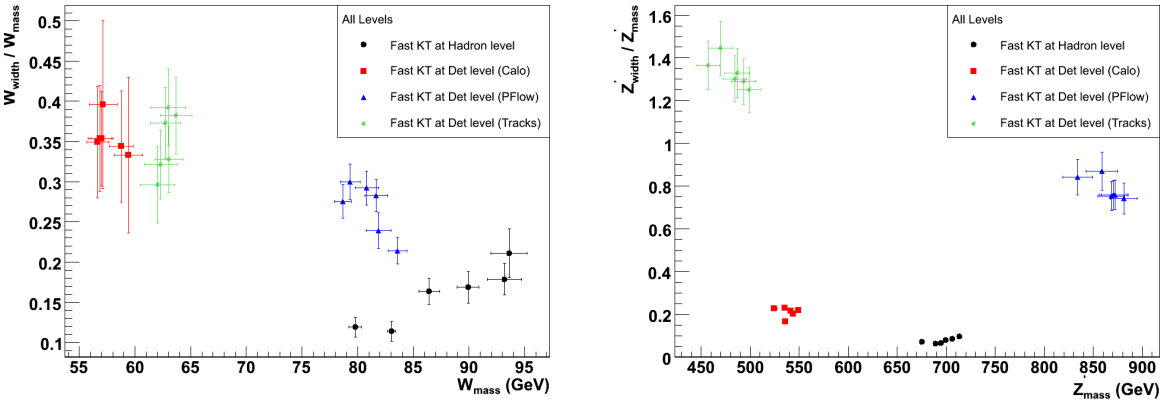

Figure 23: **Left plot:** Width / mass of the W at the hadron and detector level for different inputs and different parameters of the k_t jet-finding algorithm. **Right plot:** Width / mass of the Z$'$ at the hadron and detector level for different inputs and different parameters of the k_t jet-finding algorithm.

A global matching was performed between the hard jets and the two b-quarks from the generator level partons by minimizing the distance between the jets and the partons. If two jets were tagged as b-jets, the other two hardest jets were considered to be the light-quark jets. Furthermore both b-jets and light-jet candidates were required to be within 0.2 units in (η, ϕ). The tagged light-quark jets are used for the reconstruction of the W mass.

5.4 Reconstruction of Mass and Width of the Bosons

After di-jets were selected with the procedure explained in the previous section, the invariant mass of the di-jet system was constructed and fitted with a Breit-Wigner function:

$$p(x) = \frac{1}{2\pi} \cdot \frac{\Gamma}{(x-M)^2 + \Gamma^2/4} \,, \tag{10}$$

where x is the variable (dijet mass), M is the value of x at the maximum of the distribution and Γ the full width at half maximum of the distribution. M and Γ correspond to the mass and the width of the particle decay. One example of such a fit is shown in Fig. 20 for the case of the W boson decay at the parton level, using the partons into which the W decays, and at hadron level using the hadron level jets with the $\boldsymbol{k_t}$ algorithm with radius parameter $D = 0.4$.

The reconstructed dijet mass for Z decays at the detector level for calorimeter, track, and PFLOW $\boldsymbol{k_t}$ jets with radius parameter $D = 0.6$ is shown in Fig. 21. Also shown are the calorimeter jets where the standard Monte Carlo jet energy scale (MCJet) correction has been applied. The vertical line on the plots shows the mass of the Z boson. The mass of the Z as reconstructed with calorimeter jets is below the generated mass. This is due to the fact that no JES calibration has been applied yet. When the standard JES calibration is applied a larger value for the mass than the real one is reconstructed. The overcorrection is due to the fact that the correction is calculated for jets initiated largely by gluons, while the Z decays into two quarks. This is demonstrated in Fig. 22, where the Z decays are used for calculating quark-JES corrections that are then applied to W decays, which are also quark initiated. One can clearly see that the points with the quark-JES correction (downwards triangles with label quarks) have a peak very close to the real W mass while the points with standard corrections (upwards triangles with label CorCalo) are above the real W mass. The PFLOW jets give a reconstruction of the mass closer than calo jets to the generator level one. Track jets have a better calibration of the energy of the input particles but only 2/3 of the particles are charged, while 1/3 are neutral, on average. Therefore, track jets also give a reconstructed mass below the real value.

The ratio of the reconstructed width / mass of the Z at the detector level is shown in Fig. 21 (right). The points of the same shape (and color) represent the same input but for different D-parameters. The uncalibrated calorimeter level jets underestimate the real mass of the Z. The same holds true for the track jets, but if one takes into account that only 2/3 of the particles are charged and multiplies the reconstructed mass with track jets by 3/2 the value comes very close to the real one. Also PFLOW jets underestimate the Z mass but are much closer to it than calorimeter jets, which means that an eventual MC correction for PFLOW jets will be much smaller. The ratio values are similar for all three inputs. Also the results from generator level plots are shown. Of course they have a much better (smaller) width / mass ratio and their values are very close to the real one but a spread is observed in the reconstructed mass as a function of the D-parameter. The larger the D-parameter the more the Z mass is over-estimated. The same effect has been observed using other jet algorithms (for example, MidPoint Cone). This spread at the generator level is due to the underlying event (UE) contribution. The UE leads to production of particles with small transverse momenta which are clustered in the jets leading to the increase of the reconstructed jet transverse energy. Of course, the wider the jet (the larger the D-parameter) the larger will such an effect be. The same effect is not observed at the detector level, where the spread is smaller, because most of the (charged) low energy particles are swept away by the very large (4 T) magnetic field of the CMS solenoid and never reach the surface of ECAL. Also the calorimeter jets are reconstructed from towers which have minimum thresholds applied which further reduces the contribution of low energy particles. The subtraction of the UE at the generator level in order to correct for this effect is shown in the next section.

The results for W decays are very similar to those of Z decays. The summary of those results is shown in Fig. 23. As far as the track jets are concerned, the naive multiplication of the reconstructed mass by 3/2 does not seem to work here as in the case of the Z. The topology of the event in the W case is not as clean and not as inclusive as in the case of the Z.

The conclusions of the study for $Z^{'}$ hadronic decays are similar to those of the Z. The difference for $Z^{'}$ is that the PFLOW and tracking algorithms as of CMSSW1.3.1 do not seem to perform as well (an optimization of the PFLOW algorithms is ongoing in more recent CMSSW versions). The summary plot of width / mass for $Z^{'}$ is shown in Fig. 23.

6 Pileup and Underlying Event Subtraction

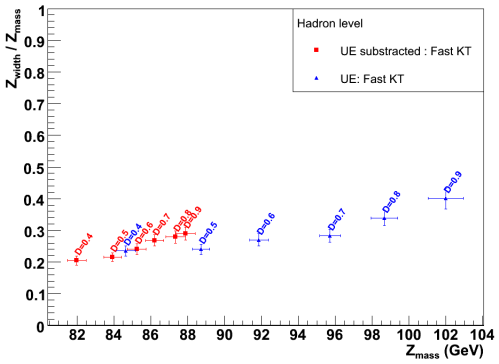

Figure 24: Width / mass of the Z at the hadron level also with application of the underlying event subtraction

A procedure for an event-by-event subtraction of pile-up (PU) and underlying event (UE) background is implemented in the FastJet package. The subtraction procedure relies on the infra-red safety of the k_t algorithm. The area of each jet in the event is calculated by filling the whole (η, ϕ) space with ghost particles of small energy and then rerunning the clustering. Afterwards the number of these ghost particles clustered in each of the newly found jets are used to calculate the area of the (original) jets. The density of the transverse momentum is calculated for each jet as p_T/A, where p_T is the transverse momentum of the jet and A its area. The median of the densities of all jets times the jet area is calculated and subtracted from the jets above a certain p_T cut. For the background subtraction in this study the following settings were used: $\eta_{max} = 6$, active area repeats = 1, ghost area = 0.01. A detailed explanation of the meaning of these parameters and the subtraction procedure is given in [5].

Both the underlying event and the pile-up interactions give approximately a constant contribution to the p_T of the jets in (η, ϕ). The PU backgrounds depends on the luminosity while the UE contribution depends on the hardness of interaction. The FastJet package gives the possibility to perform more complicated (non-constant) background fits and further studies in this direction are ongoing both for proton-proton and for heavy ion collisions. The subtraction procedure is tested here both for the subtraction of UE and of PU. In the case of PU also the UE is subtracted at the same time as the procedure makes no distinction between two constant backgrounds.

As shown in the previous section, the UE will cause an artificial over-estimation of the reconstructed Z mass in hadronic decays at the generator level. This effect can be corrected for if we apply the background subtraction. This is shown in Fig. 24 where the points with the subtracted UE have a much smaller spread than the points with the contribution of UE.

The subtration of the PU contribution is shown in Fig. 25. The reconstructed Z mass without and with the inclusion of PU is shown together with the reconstruction that performs the background subtraction. On the left plot the Z invariant mass reconstructed from calorimeter jets with D=0.7 is shown. The application of the background subtraction helps to correct the mass back to the no pile-up case and partially improves the width of the Z. The

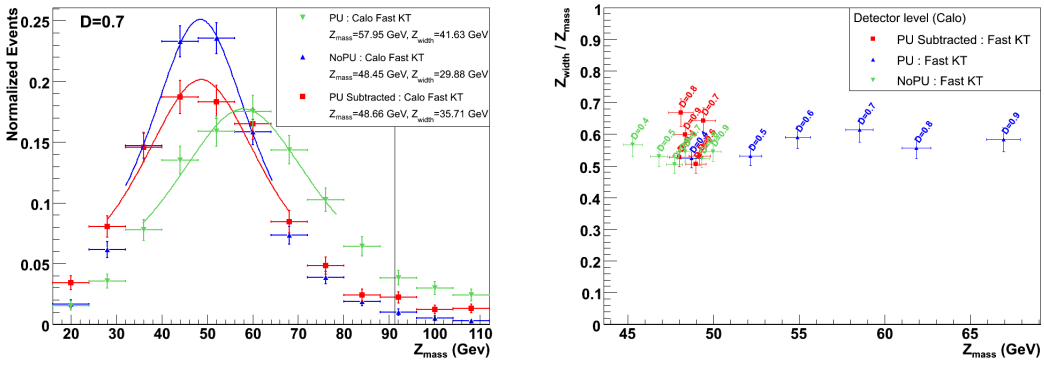

Figure 25: **Left plot:** Mass of the Z at the detector level for calo k_t jets with D=0.7 with addition of pile-up and application of the pile-up subtraction. **Right plot:** Width / mass of the Z at the detector level for calo k_t jets of different D-parameter with addition of pile-up and application of the pile-up subtraction.

absolute value of the mass peak is lower than the real Z mass since the calorimeter jets are uncorrected, as already explained. On the right plot of Fig. 25 the effect can be seen for different D-parameters. Clearly, the wider the jet the larger is the effect of PU. The subtraction procedure seems to work nicely for all of them correcting back to almost the same value of the mass. The PU subtraction has been observed to work the same way for the W and Z$'$ cases.

7 Conclusions

First studies of the performance of the k_t jet algorithm in CMSSW have been performed. The fast implementation of the k_t from the FastJet package has been validated and will be the default k_t algorithm used in CMS in the future. This implementation was measured to be approximately 40 times faster than the standard implementation of the k_t when run on calorimeter towers. The produced jet collections are, as expected, identical as long as care is taken in order to avoid ill-defined input four-vectors with negative mass squares. The time spent per event of the fast implementation is similar to the one of the simple iterative cone jet finder.

The jet reconstruction efficiency as a funtion of p_{T} and η has been studied. The efficiency turns out to be larger with a D parameter of 0.6 compared to 1. The differences in efficiency observed in the different calorimeter regions is reduced as p_{T} increases. The jet response has been studied and is similar with a D parameter of 0.6 compared to 1. The jet energy resolution slightly improves with a smaller D parameter value of 0.6 and the position resolution improves substantially with a D parameter of 0.6 compared to 1.

Differential jet shapes have been studied. It has been seen that 90% of the jet energy of the calorimeter level jets is confined to $r < 0.27, 0.37$ and 0.55 for values of $D = 0.4, 0.6$ and 1.0, respectively, where r is the distance between the jet position and its constituents.

Hadronic decays of Z/Z$'$ and W have been used as a testbench for the performance of the k_t jet finder. The k_t algorithm performs well in the reconstruction of these hadronic decays. A general preference for lower D-parameters is observed as narrow jets are less influenced from underlying event and pile up contributions. The hadronization factors might be larger for these narrower jets, though. Standard CMS JES corrections overcorrect the reconstructed mass of Z and W with respect to the true (generator level) value. This is due to the fact that standard corrections are calculated for jets initiated largely by gluons, while Z and W decay into two quarks. If the jets coming from Z decays are used to derive new quark-JES corrections and these corrections are subsequently applied to W decays a proper recontruction of the W mass is obtained. The event-by-event background subtraction method provided in the FastJet package has been successfully tested and has shown a good performance in correcting for effects of both underlying event and pileup. These studies about the PU / UE subtraction are preliminary and show the potential of the method. Further detailed studies are ongoing.

References

[1] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, *"Longitudinally invariant k_t clustering algorithms for hadron hadron collisions"*, Nucl.Phys.B **406**:187-224 (1993). .

[2] **ORCA**, http://cmsdoc.cern.ch/ORCA

[3] **CMSSW**, https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideFrameWork

[4] J. M. Butterworth, J. P. Couchman, B. E. Cox, B. M. Waugh, *"KtJet: A C++ implementation of the k_t clustering algorithm"*, Comp. Phys. Comm. vol 153/1 85-96 (2003).

[5] M. Cacciari, G. P. Salam, *"Dispelling the N^3 myth for the Kt jet-finder"*, Phys.Lett. B**641** 57-61 (2006).

[6] R. Demina et al., *Calorimeter Cell Energy Thresholds for Jet Reconstruction in CMS*, CMS NOTE-2006/020 (2006).

[7] http://proj-clhep.web.cern.ch/proj-clhep

[8] T. Sjostrand, S. Mrenna and P. Skands, *"PYTHIA 6.4 physics and manual"*, JHEP **0605**, 026 (2006).

[9] M. Vazquez Acosta et al., *Jet and MET Performance in CMSSW 1.2.0*, CMS IN-2007/053 (2007).

[10] F. Maltoni and T. Stelzer, *"MadEvent: Automatic event generation with MadGraph"*, JHEP **0302**, 027 (2003).